Accordingly, such constraints can be represented by coordinates in three dimensions, for example, as having a certain position, or range of positions, along x, y, and z coordinates (*i.e.*, a "coordinate set"). Alternatively, a geometric or tertiary constraint can be represented as a distance, or range of distances, between a particular atom (or pseudoatom, group of atoms, *etc.*) and another atom (or pseudoatom, group of atoms, *etc.*). Tertiary constraints can also be represented by various types of angles, including the angle of bonds (particularly covalent bonds, *e.g.*, φ bonds and ψ bonds) between atoms in an amino acid residue, between atoms in different amino acid residues, and between atoms in an amino acid residue of a protein and another molecule, *e.g.*, a ligand, with ranges for each angle being preferred.

A "conformational constraint" or "secondary constraint" refers to the presence of a particular protein conformation, for example, an α-helix, parallel and antiparallel β strands, leucine zipper, zinc finger, *etc.* in which an amino acid residue, or group of residues, is located. In addition, conformational or secondary constraints can include amino acid sequence information without additional structural information. As an example, "-C-X-X-C-" is a conformational constraint indicating that two cysteine residues must be separated by two other amino acid residues, the identities of each of which are irrelevant in the context of this particular constraint.

An "identity constraint" refers to a constraint that indicates the identity of a particular amino acid residue at a particular amino acid position in a protein. Typically, an amino acid position is determined by counting from the amino-terminal residue of the protein up to and including the residue in question. As those in the art will appreciate, comparison between related proteins may reveal that the identity of a particular amino acid residue at a given amino acid position in a protein is not entirely conserved, *i.e.*, different amino acid residues may be present at a particular amino acid position in related proteins, or even in allelic or other variants of the same protein.

To "relax" a constraint refers to the inclusion of a user-defined variance therein. The degree of relaxation will depend on the particular constraint and its application.

As those in the art are aware, protein structures can be of different quality. Presently, the highest quality determination methods are experimental structure prediction methods based on x-ray crystallography and/or NMR spectroscopy. In x-ray crystallography, "high resolution" structures are those wherein atomic positions are determined at a resolution of about 2 Å or less, and enable the determination of the three-dimensional positioning of each atom (or at least each non-hydrogen atom) of a protein. "Medium resolution" structures are those wherein atomic positioning is determined at about the 2-4 Å level, while "low resolution" structures are those wherein the atomic positioning is determined in about the 4-8 Å range. Herein, protein structures that have been determined by x-ray crystallography or NMR may be referred to as "experimental structures," as compared to those determined by computational methods, *i.e.*, derived from the application of one or more computer algorithms to a primary amino acid sequence to predict protein structure.

As alluded to above, protein structures can also be determined entirely by computational methods, including, but not limited to, homology modeling, threading, and *ab initio* methods. Often, models produced by such computational methods are "reduced" models. A "reduced model" refers to a three-dimensional structural model of a protein wherein fewer than all heavy atoms (*e.g.*, carbon, oxygen, nitrogen, and sulfur atoms) of the protein are represented. For example, a reduced model might consist of just the α-carbon atoms of the protein, with each amino acid connected to the subsequent amino acid by a virtual bond. In one embodiment, reduced models are those comprised only of side chain centers of mass. As will be appreciated by those in the art, more detailed model structures of a protein can be assembled from a reduced model. For example, a reduced model comprised only of amino acid residue side chain centers of mass implicitly specifies the location of the atoms comprising the side chain, as well the position of the

peptide backbone. Accordingly, whatever greater level of atomic detail is required, if any, for the particular application can be added to a reduced model, and it is understood that once a protein structure based on a reduced model has been generated, all or a portion of it may be further refined to include additional predicted detail, up to including all atom positions.

Computational methods usually produce lower quality structures than experimental methods, and the models produced by computational methods are often called "inexact models." While not necessary in order to practice the instant methods, the precision of these predicted models can be determined using a benchmark set of proteins whose structures are already known. For example, the predicted model can be compared to a corresponding experimentally determined structure. The difference between the predicted model and the experimentally determined structure is quantified via a measure called "root mean square deviation" (RMSD). A model having an RMSD of about 2.0 Å or less as compared to a corresponding experimentally determined structure is considered "high quality". Frequently, predicted models have an RMSD of about 2.0 Å to about 6.0 Å when compared to one or more experimentally determined structures, and are called "inexact models". As those in the art will appreciate, RMSDs can also be determined for one or more atomic positions when two or experimental structures have been generated for the same protein.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Illustration of the protein chain representation. (A) For a short expanded fragment and (B) for a helical fragment. The solid circles correspond to explicitly simulated side chain centers of mass. The open circles indicate the expected positions of the α-carbons.